COMMONSENSE MINING AS KNOWLEDGE BASE COMPLETION?
A STUDY ON THE IMPACT OF NOVELTY
Stanislaw Jastrzębski, Dzmitry Bahdanau, Seyedarian Hoseinei, Michael Noukhovitch, Yoshua Bengio, Jackie Chi Kit Cheung

Abstract
- We analyze whether knowledge base completion models can be used to mine commonsense.
- We propose "novelty" with respect to the training set as an important factor in evaluation, and use it to show a simpler model outperforms state-of-the-art.

Mining Commonsense
- Many NLP tasks require commonsense knowledge but collecting and organizing it is difficult.
- Commonsense knowledge bases (e.g., ConceptNet) represent commonsense knowledge as relational triples: ("pen", "UsedFor", "writing").

Mining as Knowledge Base Completion
- A common way of improving the coverage of knowledge bases is through knowledge base completion (KBC).
- Li et al. approached commonsense mining as a KBC task. Their method mines candidate triples from Wikipedia and ranks the triples with a KBC model in order to extend ConceptNet.

Models for KBC
- All our models take \((h, r, t)\) triples as inputs, where \(h\) and \(t\) are sequences of words representing concepts and \(r\) is a relation. We embed \(h\) and \(t\) by embedding each word and summing over the sequence to get \(h\) and \(t\) and we embed the relation to get \(r\).

- Bilinear: \(h^r\)\(M_r\)\(t\).
- DNN: \(W_2\)\(W_1\)\((h, r, t)\) + \(b_1\) + \(b_2\).
- Factorized: \(\alpha(Ah + b_1, Bh + b_2) + \beta(Ar + b_1, Br + b_2) + \gamma(At + b_1, Tt + b_2)\).
- Prototypical: \(\beta(At + b_1, Br + b_2) + \gamma(At + b_1, Tt + b_2)\).

ConceptNet and Wikipedia Setup
- Models are trained using 100k triples from ConceptNet5 that were extracted from the OMCS corpus.

- ConceptNet5 - Completion task considers two ways to split the dataset: a random split, and confidence-based split using triples with the highest confidence scores as a test set.

ConceptNet and Wikipedia Results Mismatch
- We find that the knowledge base completion task is a poor indicator of performance on Wikipedia.

- Factorized and Prototypical models achieve a similar or worse score compared to DNN on the KBC task.

Novelty Explains the Mismatch
- In Category 1 we find ("egg", "HoA", "food"), which has a close analog in the training set: ("egg", "HoA", "type of food").

- In Category 3 ("different relation and word") we find ("floor", "UsedFor", "walk on"), which has a corresponding triple in the training set ("floor", "UsedFor", "stand on").

Acknowledgments: SJ was supported by Grant No. DI 2014/16644 from Ministry of Science and Higher Education, Poland. Work at MILA was funded by NSERC, CIFAR, and Canada Research Chairs.